ELECTRICITY LOAD PROFILE CLASSIFICATION USING FUZZY C-MEANS METHOD

Iswan Prahastono
PT. PLN Indonesia, University
of Abertay Dundee, UK,
0604721@abertay.ac.uk

Dr. David J King
University of Abertay
Dundee, UK
D.king@abertay.ac.uk

Dr. C.S. Ozveren
University of Abertay
Dundee, UK,
c.s.ozveren@abertay.ac.uk

Prof. D.Bradley
University of Abertay
Dundee, UK
D.Bradley@abertay.ac.uk

Abstract- **This paper presents the Fuzzy C-Means (FCM) clustering method. The FCM technique assigns a degree of membership for each data set to several clusters, thus offering the opportunity to deal with load profiles that could belong to more than one group at the same time. The FCM algorithm is based on minimising a c-means objective function to determine an optimal classification.**

The simulation of FCM was carried out using actual sample data from Indonesia and the results are presented. Some validity index measurements was carried out to estimate the compactness of the resulting clusters or to find the optimal number of clusters for a data set.

I. INTRODUCTION

New Indonesian government regulations in the electricity sector will be introduced in the near future, which will allow regions to develop their own electricity tariff structures. At present, tariffs are classified based on data collected over several years. This historical database of customer load demand profiles is used to group customers by variations in load pattern, such as industrial, business, public services and residential loads. Considering the geographic layout of Indonesia, which covers thousands of islands with differing socio-economic circumstances, the most appropriate method to apply may vary between regions or islands. Hierarchical techniques might be applied to develop a new tariff structure in a new region. A hard clustering method could be applied in some regions where the types of customers are known, so the number of classes can be predicted. This paper will concentrate on the Fuzzy C-means (FCM) classification method for clustering electricity load profiles and will show why this method could be appropriate for other conditions not mention above and applicable in Indonesia.

II. THEORY OF FUZZY C-MEANS CLUSTERING

The Fuzzy C-means (FCM) method is a clustering technique wherein each data point belongs to every cluster to some degree that is specified by a membership grade. The procedure is similar to standard K-means [1], the difference is that each data set has a degree of membership to each initial cluster [2], i.e. each data set belongs to all clusters to some degree. The FCM process does not create boundaries between data sets for the first iteration, because the clustering process involves all data. The iteration is based on minimizing an objective function that represents the distance from any data point of the load profile to a cluster centre point weighted by their membership grade. The boundaries will automatically evolve when the clustering process is completed [3].

The procedure starts with determining the number of clusters and guessing the cluster centre point (most likely incorrect), which is intended to mark the mean location of each cluster, then assigning every data set a membership grade for each cluster. The next step is to update each cluster centre point and membership grade iteratively until the position of the centre point is stable. In this step, the cluster centre point moves iteratively to the correct position within the data sets. This iteration is based on minimizing an objective function that represents the distance from all data to a cluster centre, weighted by a degree of membership.

The FCM can be applied to data that is quantitative (numerical), qualitative (categorical), or a combination of both. The load profile data can be arranged as a matrix where each row contains load profile data corresponding to

observed variations of customer load over time and the column corresponds to the number of evaluated customers.

The FCM method allows objects to belong to several of the determined clusters simultaneously, with different degrees of membership. The data is partitioned into fuzzy subsets, so the objects on the boundaries between several classes are not forced to fully belong to one group, but rather are assigned membership degrees between 0 and 1 indicating their partial membership. Thus, the load profile cluster centre is the mean of all data points in the same observation, weighted by their degree of belonging to the cluster [4].
The distance between the cluster centre and each data point can be calculated by using Euclidean Distance.

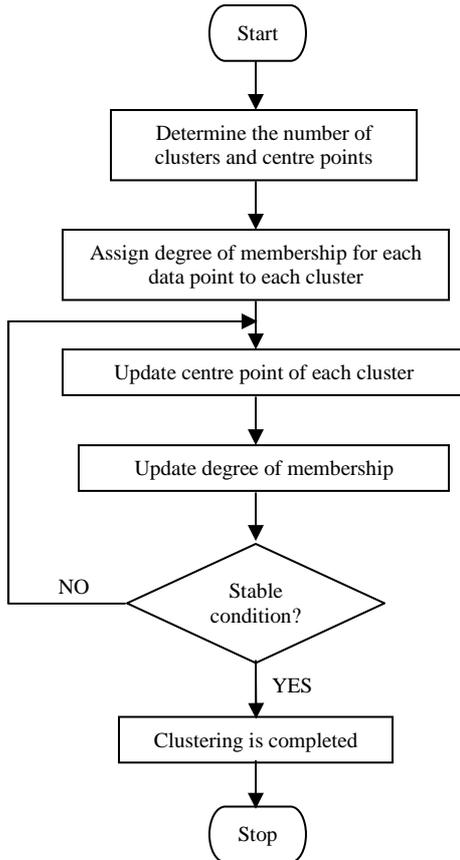The flowchart for this procedure is shown in figure 1 below.

Fig. 1.   Flowchart of Fuzzy C-Means Clustering

III. SIMULATION

The specific fuzzy function in the Matlab Fuzzy Logic toolbox has been applied for all sequence simulations of Fuzzy C-Means clustering.

The simulation uses hourly load data taken from different types of electricity consumers' in Indonesia, which was collected from Automatic Meter Reading (AMR) on the same time and date. The sequence of pre-processing and simulation using the data is described below.

A. Normalising Data

All collected consumers' data from AMR must be normalised, by dividing by the maximum or the largest value of the data, in this case the maximum load consumed (kW).

$$l_{(n,h-norm)} = \frac{l_{(n,h)}}{l_{h-max}} \tag{1}$$

Where :

n : indicates the n-th load profile data set; $n = 1, ... , N$ (N = total number of load profiles).

h : indicates the h-th data point in a load profile data set; $h = 1, ... , H$ (H = total number of data points).

$l_{(n,h-norm)}$: normalised value of data.
$l_{(n,h)}$: actual value of data.
l_{h-max} : the maximum value of data.

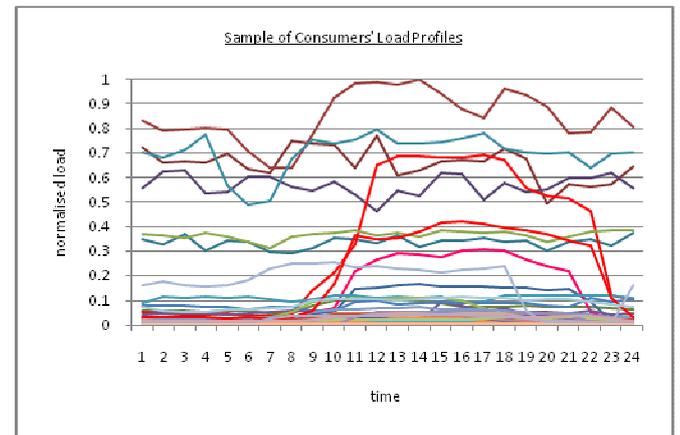The normalised sample consumers' load profiles data collected from Indonesia is shown in figure 2 below.

Fig. 2.   Sample of Load Profiles

B. Simulation Process

The simulation sequence of the Fuzzy C-means method can be generalised as follows :

1. Determine the number of expected clusters (C) and random position of cluster centres.
2. Assign the degree of membership for each observation data point for all load profile data sets to each cluster centre within the same time point, thus they have a degree of coefficient or partition (μ) and the sum of these is defined to be 1 :

$$\forall l_{(i,h)} \quad \sum_{c=1}^{C} \mu_c(l_{(i,h)}) = 1 \tag{2}$$

Where :

$l_{(i,h)}$: indicates the h-th observation data point in a set of load profiles $l_{(i)}$.

μ_c : indicates the structure of partition matrix for each cluster centre (c).

The load profile cluster centre is the mean of all data points in the same observation, weighted by their degree of belonging to that cluster :

$$l_{(c,h)} = \frac{\sum\limits_{c=1}^{C} \mu_c(l_{(i,h)})^m l_{(i,h)}}{\sum\limits_{c=1}^{C} \mu_c(l_{(i,h)})}$$

(3)

Where :

$l_{(c,h)}$: indicates the *h-th* observation data point in a load profile cluster centre $l_{(c)}$.

$\mu_c(l_{(i,h)})$: indicates the partition matrix for $l_{(i,h)}$.

m : weighting exponent.

3. Calculate the distance between load profile data sets and each cluster centre by using Euclidean distance as described in equation (*2*).

4. Repeat steps 2 and 3 iteratively until convergence is achieved (the iteration is based on minimising an objective function that represents the distance from all data to a cluster centre, weighted by a degree of membership). This means that the distance between each load profile data set to the nearest centre point is stable.

C. Simulation Results

The results of simulations using several values of the weighting exponent, which indicates differing degrees of membership, are shown in figures 3-6 below.
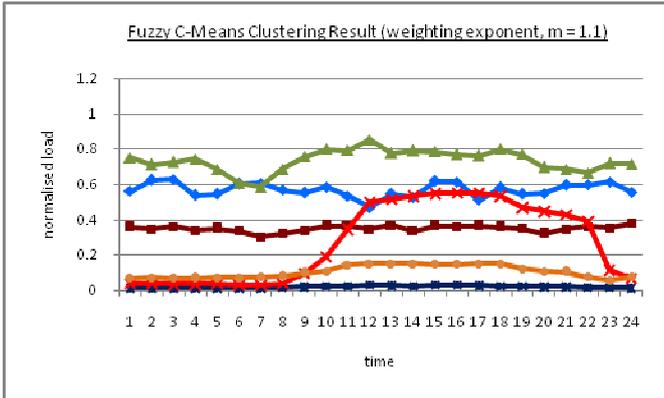

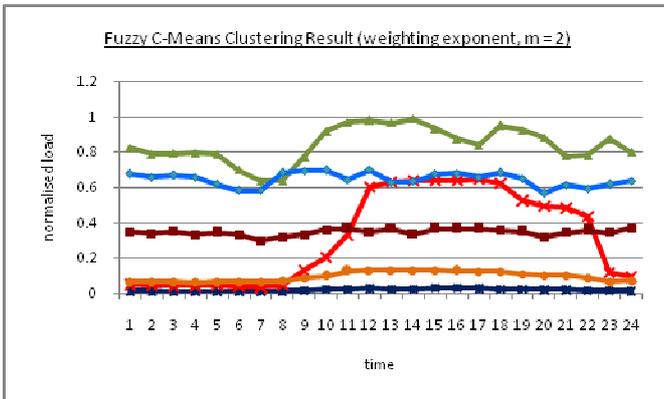Fig. 3.   Fuzzy C-Means Clustering Result (m=1.1)

Fig. 4.   Fuzzy C-Means Clustering Result (m=2)
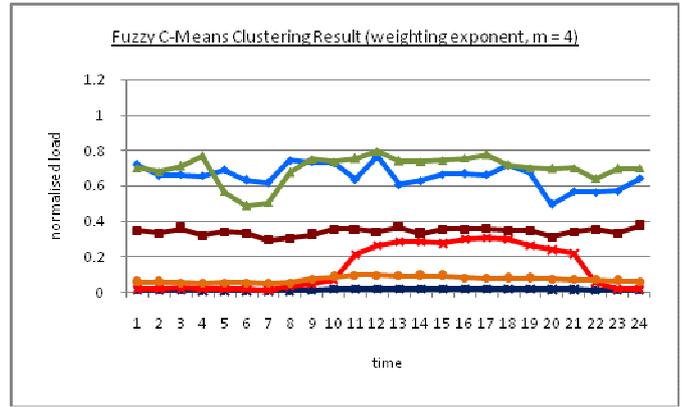

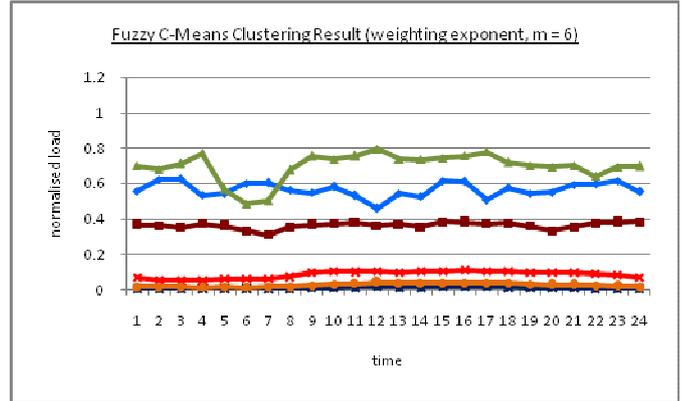Fig. 5.   Fuzzy C-Means Clustering Result (m=4)

Fig. 6.   FCM Clustering Result (m=6)

The purpose of the simulation is to minimise the objective function that represents the distance from all data to the cluster centre, weighted by a degree of membership. Figure 7 below shows an example, from one simulation, of the value of the objective function changing until stable during the iteration process.
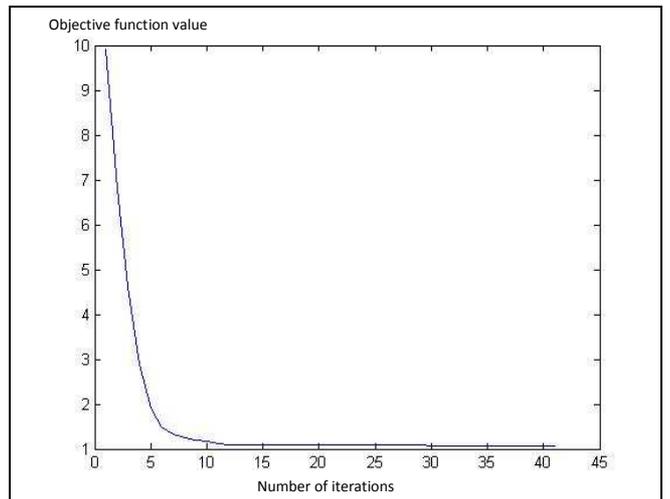

Fig. 7.   Sample of Objective Function Value

IV. RESULTS ANALYSIS

The different weighting exponent values will produce different sets of load profile groups. Increasing the weighting exponent value causes the load profiles to become more flat,

because the degree of membership for each data eventually belong to all clusters equally.

Compactness of the clustering result can be measured using Dunn Index method [5]

$$DI(c) = min_{i \in c} \left\{ min_{j \in c, i \neq j} \left(\frac{d(c_i, c_j)}{max_{k \in c} dia(c_k)} \right) \right\}$$

(4)

Where, $d(c_i, c_j)$ is the dissimilarity function between two clusters c_i and c_j defined as

$$d(c_i, c_j) = min_{x \in c_i, y \in c_j} d(x, y)$$

(5)

And $dia(c)$ is the diameter of a cluster, which may be considered as a measure of dispersion of the clusters. The diameter of a cluster 'c' can be defined as follows:

$$dia(c) = max_{x,y \in c} d(x, y)$$

(6)

All the equations above show that a higher value of Dunn Index indicates compactness and good separation of the clusters as shown in table 1 below.

TABLE I
DUNN INDEX RESULTS

Dunn Index Results with different weighting exponent (m)			
m = 1.1	m = 2	m = 4	m = 6
0.1580	0.0845	0.0290	0.0256

The Dunn Index does not exhibit any trend with respect to number of clusters, thus, the optimum number of clusters for a certain value of weighting exponent can be identified only by trial and error. Figure 8 below shows an example of Dunn Index value for different numbers of resulting clusters.
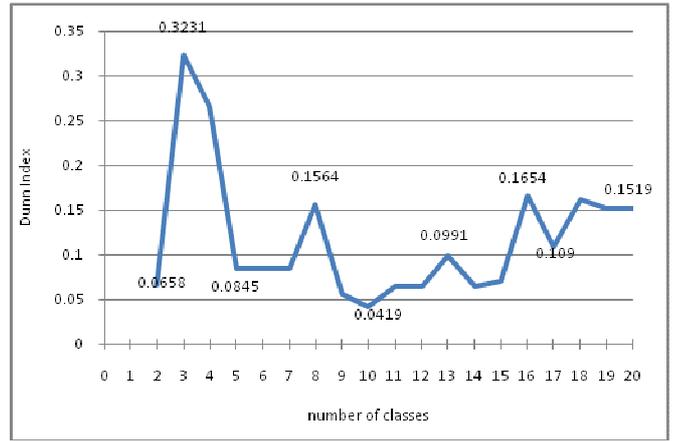

Fig. 8.   Dunn Index (DI) Value

Overlapping between clusters can be measured using the Partition Index (PC), which has the following formula :

$$PC(c) = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{N} (u_{ij})^2$$

(7)

where μ_{ij} is the degree of membership of data point j to cluster i.

The optimum number of clusters is indicated by the maximum value of PC. Figure 9 below shows an example of the value of PC index in comparison to the number of clusters produced.

Fig. 9.   Partition Coefficient Value

The fuzzyness of the cluster partition can be measured using Classification Entropy (CE) which has the following formula:

$$CE(c) = -\frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{N} \mu_{ij} \log(\mu_{ij})$$

(8)

which is similar to the PC index, the highest value indicates the optimum number of clusters. Figure 10 below shows an example of the value of CE for various number of clusters.
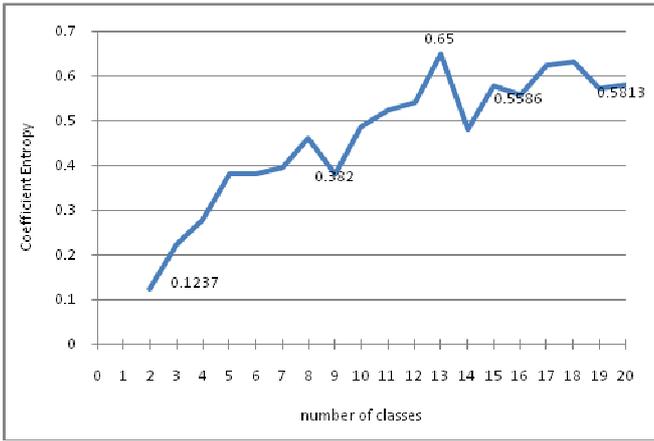
Fig. 10. Coefficient Entropy Value

V. DISCUSSIONS

Choosing a small value of weighting exponent (close to one) is useful in performing clustering for consumers' load profiles, that have very contrasting curve shapes. The DI shows the highest value, in the simulation, for the weighting exponent, $m = 1.1$ (the smallest). This is understandable, because it is very close to crisp clustering, where the clusters are well separated. Figures 11, 12, and 13 below show the simulation results for Hierarchical, K-Means and FCM methods with the same number of resultant clusters.
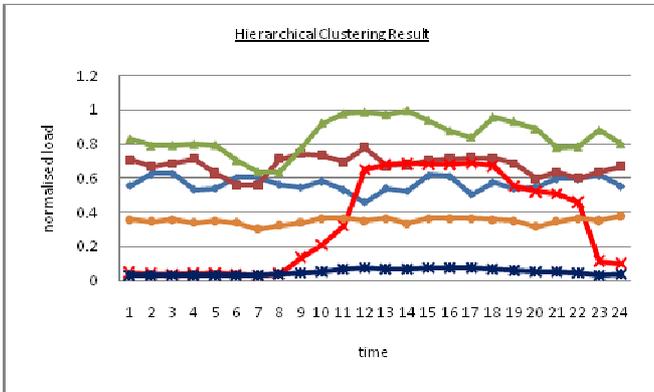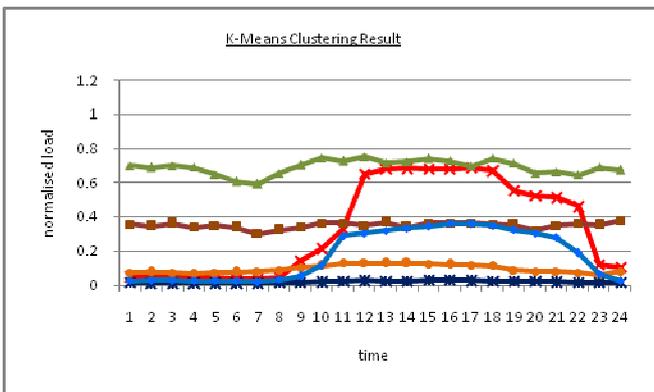

Fig. 11. Hierarchical Clustering Result
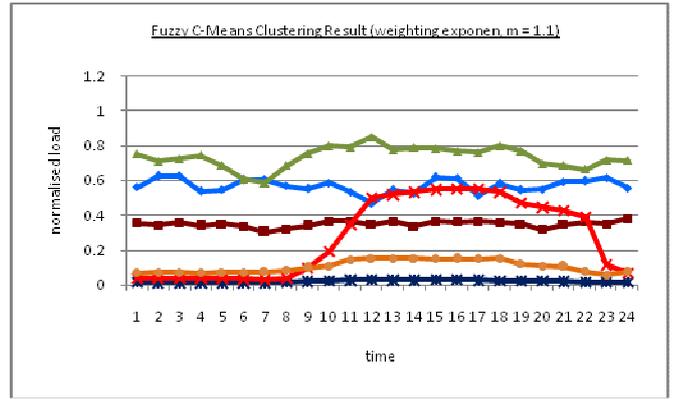

Fig. 12. K-Means Clustering Result

Fig. 13. Fuzzy C-Means Clustering Result (m=1.1)

The figures above shown that those techniques can separate and correctly cluster load profiles with a wide fluctuation of load value (indicated in x red). So, the FCM with the small value of weighting exponent is able to separate the clusters equally as well as the hard clustering method, as shown in TABLE .

The FCM techniques could be applied in Indonesia, where a particular region or area wants to perform new clusters of load profiles for the purpose of developing a new tariff design. The consumers will be initially classified into major groups that shown significant differences between load profiles, such as industrial, business, public facilities and residential.

The number of groups can be evaluated as follows :

- The Dunn Index can be used, where new groups are required, hence the compactness and separation between clusters is the priority.
- The Partition Index can be used to perform sub-classes within the major groups where the overlapping between clusters becomes the priority consideration during the grouping process.
- The Coefficient Entropy can be used to perform groupings where the fuzzyness of the clusters becomes the priority consideration during the grouping process.

If it is necessary to perform clusters from consumers' load profiles which have similar patterns, this can be carried out by choosing a higher value of weighting exponent. This technique, in Indonesia, will be useful to develop sub-classes within the major groups based on installed capacity. For example, in the major residential group, there are several sub-classes with different installed capacity.

VI. CONCLUSION

FCM has greater flexibility than other load profiles clustering method by applying the different values of weighting exponent, wide variety of groupings can be produced which depend on the users' requirements. This method could become very useful in Indonesia for producing new tariff structures in the various differing regions.

ACKNOWLEDGMENT

REFERENCES

[1]. Prahastono, I., King, D., and Ozveren, C., S., *A review of electricity load profila classification method*, Accepted paper of UPEC Conference, Brighton, 2007.

[2]. Chicco, G.. Napoli, R., Piglione, F., *Application of clustering algorithms and self organising maps to classify electricity customers*, Power Tech Conference Proceedings, 2003 IEEE Bologna Volume 1, 23-26 June 2003, pp.7

[3]. Tai Wai Cheng, Dmitry B. Goldgof, Lawrence O. Hall, *Fast Fuzzy* clustering, ISAAC: 6th International Symposium on Algorithms and Computation, 1995.

[4]. Almeida, R., J., and Sousa, J., M., C., *Comparison of fuzzy clustering algorithms for classification*, International Sympposium on Evolving Fuzzy Systems , IEEE, September, 2006.

[5]. Chicco, G.. Napoli, R., Postolache, P., Scutariu, M., and Toader, C., *Customer characterisation options for improving the tariff order*, IEEE Transaction on Power Systems, 2003, Volume 18, No.1, February 2003.